

Homogeneity of Cluster Ensembles

Brijnesh J. Jain

Technische Universität Berlin, Germany

e-mail: brijnesh.jain@gmail.com

The expectation and the mean of partitions generated by a cluster ensemble are not unique in general. This issue poses challenges in statistical inference and cluster stability. In this contribution, we state sufficient conditions for uniqueness of expectation and mean. The proposed conditions show that a unique mean is neither exceptional nor generic. To cope with this issue, we introduce homogeneity as a measure of how likely is a unique mean for a sample of partitions. We show that homogeneity is related to cluster stability. This result points to a possible conflict between cluster stability and diversity in consensus clustering. To assess homogeneity in a practical setting, we propose an efficient way to compute a lower bound of homogeneity. Empirical results using the k-means algorithm suggest that uniqueness of the mean partition is not exceptional for real-world data. Moreover, for samples of high homogeneity, uniqueness can be enforced by increasing the number of data points or by removing outlier partitions. In a broader context, this contribution can be placed as a further step towards a statistical theory of partitions.

Contents

1. Introduction	3
2. Partition Spaces	4
2.1. Partitions	4
2.2. Orbit Spaces	5
2.3. Intrinsic Metric	5
3. Homogeneity of a Sample	6
3.1. Fréchet Functions	6
3.2. Conditions of Uniqueness	6
3.3. Asymmetric Partitions	7
3.4. Homogeneity	9
3.5. Clustering Stability	10
3.5.1. Clustering Instability	10
3.5.2. Homogeneity vs. Stability	11
3.5.3. Stability vs. Diversity	12
4. Experiments	13
4.1. Experiments on Synthetic Data	13
4.1.1. Results on G4 Datasets	14
4.1.2. Results on U-Shapes and Gaussians	18
4.2. Experiments on UCI Datasets	20
5. Conclusion	21
A. Preliminaries	22
A.1. Notations	22
A.2. Dirichlet Fundamental Domains	23
A.3. Cross Sections	24
B. Proofs	24
B.1. Proof of Theorem 3.1	24
B.2. Proof of Prop. 3.2	26
B.3. Proof of Prop. 3.3	27
B.4. Proof of Prop. 3.5	27
B.5. Proof of Equation (1)	28

1. Introduction

Clustering is a standard technique for exploratory data analysis that finds applications across different disciplines such as computer science, biology, marketing, and social science. The goal of clustering is to group a set of unlabeled data points into several clusters based on some notion of dissimilarity. Inspired by the success of classifier ensembles, consensus clustering has emerged as a research topic [9, 22]. Consensus clustering first generates several partitions of the same dataset. Then it combines the sample partitions to a single consensus partition. The assumption is that a consensus partition better fits to the hidden structure in the data than individual partitions.

One standard approach of consensus clustering combines the sample partitions to a mean partition [3, 4, 6, 7, 10, 14, 18, 20, 21]. A mean partition best summarizes the sample partitions with respect to some (dis)similarity function. In general, a mean partition is not unique. Non-uniqueness of a mean partition poses a number of challenges, including (i) comparability, (ii) consistency, (iii) asymptotic behavior, and (iv) clustering stability.

Uniqueness allows us to directly compare the performance of two clustering ensembles via their respective mean partitions. Without uniqueness, comparing two distributions of partitions based on randomly generated samples can be elusive. Moreover, under reasonable conditions, uniqueness implies strong consistency and gives rise to different versions of the law of large numbers [13, 19]. These findings indicate that without uniqueness, statistical inference based on mean partitions can hardly proceed. With regard to clustering stability, non-uniqueness of the mean partition could potentially entail instability of the clustering.

Despite being a desirable property coming along with several benefits, not much research has been devoted to uniqueness of the mean partition in consensus clustering. In particular, it is unclear under which conditions a sample has a unique mean partition. In addition, it is also unclear whether uniqueness only occurs in trivial and exceptional cases or is a feasible property of practical relevance. To approach these issues, we assume that the set of (hard and soft) partitions is endowed with an intrinsic metric induced by the Euclidean distance.

The contributions of this paper are as follows:

1. *Conditions of uniqueness.* We establish conditions of uniqueness showing that the mean partition is unique if the cluster ensemble generates sample partitions within a sufficiently small ball.
2. *Homogeneity.* We propose homogeneity as a measure of how close a sample is to having a unique mean partition. We present a lower-bound of homogeneity that can easily be computed and at the same time identifies outlier-partitions that need to be removed in order to guarantee a subsample with unique mean partition.
3. *Relationship to cluster stability.* We show that homogeneity of a sample is related to cluster stability. This result points to potentially colliding approaches in clustering: standard clustering advocates stability [11, 16] and consensus clustering

advocates diversity [5, 22, 23], which will be briefly discussed.

4. *Empirical evidence.* In experiments we assessed the homogeneity of samples obtained by the k-means algorithm applied to synthetic and real-world data. The results suggest that uniqueness of the mean partition is not exceptional and can be enforced by a larger dataset or by removing outlier-partitions if homogeneity is high.

Though uniqueness of the mean partition is not a side issue, it is still a strict property not valid for many samples. Homogeneity relaxes this strict property and gives us an alternative way to assess the performance of clusterings and cluster ensembles that goes beyond uniqueness of the mean partition. In a wider context, the results presented in this paper contribute towards a statistical theory of partitions [13].

The rest of this paper is structured as follows: Section 2 represents partitions as points of an orbit space. In Section 3, we present the theoretical contributions. Section 4 discusses experimental results. Finally, Section 5 concludes with a summary of the main results and with an outlook to further research. Proofs are delegated to the appendix.

2. Partition Spaces

To analyze partitions, we suggest a geometric representation proposed in [13]. We first show that a partition can be regarded as a point in some geometric space, called orbit space. Orbit spaces are well explored, possess a rich geometrical structure and have a natural connection to Euclidean spaces [2, 12, 17]. Then we endow orbit spaces \mathcal{P} of partitions with a distance function δ related to the Euclidean metric such that (\mathcal{P}, δ) becomes a geodesic metric space.

2.1. Partitions

Let $\mathcal{Z} = \{z_1, \dots, z_m\}$ be a set of m data points. A partition X of \mathcal{Z} with ℓ clusters $\mathcal{C}_1, \dots, \mathcal{C}_\ell$ is specified by a matrix $\mathbf{X} \in [0, 1]^{\ell \times m}$ such that $\mathbf{X}^T \mathbf{1}_\ell = \mathbf{1}_m$, where $\mathbf{1}_\ell \in \mathbb{R}^\ell$ and $\mathbf{1}_m \in \mathbb{R}^m$ are vectors of all ones.

The rows $\mathbf{x}_{k:}$ of matrix \mathbf{X} refer to the clusters \mathcal{C}_k of partition X . The columns $\mathbf{x}_{:j}$ of \mathbf{X} refer to the data points $z_j \in \mathcal{Z}$. The elements x_{kj} of matrix $\mathbf{X} = (x_{kj})$ represent the degree of membership of data point z_j to cluster \mathcal{C}_k . The constraint $\mathbf{X}^T \mathbf{1}_\ell = \mathbf{1}_m$ demands that the membership values $\mathbf{x}_{:j}$ of data point z_j across all clusters must sum to one.

By $\mathcal{P}_{\ell, m}$ we denote the set of all partitions with ℓ clusters over m data points. Since some clusters may be empty, the set $\mathcal{P}_{\ell, m}$ also contains partitions with less than ℓ clusters. Thus, we consider $\ell \leq m$ as the maximum number of clusters we encounter. If the exact numbers ℓ and m do not matter or are clear from the context, we also write \mathcal{P} for $\mathcal{P}_{\ell, m}$. A hard partition X is a partition with matrix representation $\mathbf{X} \in \{0, 1\}^{\ell \times m}$. The set $\mathcal{P}^+ \subset \mathcal{P}$ denotes the subset of all hard partitions.

2.2. Orbit Spaces

We define the *representation space* \mathcal{X} of the set $\mathcal{P} = \mathcal{P}_{\ell,m}$ of partitions by

$$\mathcal{X} = \{\mathbf{X} \in [0, 1]^{\ell \times m} : \mathbf{X}^T \mathbf{1}_\ell = \mathbf{1}_m\}.$$

Then we have a natural projection

$$\pi : \mathcal{X} \rightarrow \mathcal{P}, \quad \mathbf{X} \mapsto X = \pi(\mathbf{X})$$

that sends matrices \mathbf{X} to partitions X they represent. The map π conveys two properties: (1) each partition can be represented by at least one matrix, and (2) a partition may have several matrix representations.

Suppose that matrix $\mathbf{X} \in \mathcal{X}$ represents a partition $X \in \mathcal{P}$. The subset of all matrices representing X forms an equivalence class $[\mathbf{X}]$ that can be obtained by permuting the rows of matrix \mathbf{X} in all possible ways. The equivalence class of \mathbf{X} is of the form

$$[\mathbf{X}] = \{\mathbf{P}\mathbf{X} : \mathbf{P} \in \Pi\},$$

where Π is the group of all $(\ell \times \ell)$ -permutation matrices. The *orbit space of partitions* is the set

$$\mathcal{X}/\Pi = \{[\mathbf{X}] : \mathbf{X} \in \mathcal{X}\}.$$

Informally, the orbit space consists of all equivalence classes $[\mathbf{X}]$, we can construct as described above. Mathematically, the orbit space \mathcal{X}/Π is the quotient space obtained by the action of the permutation group Π on the set \mathcal{X} . The equivalence classes $[\mathbf{X}]$ are the orbits of \mathbf{X} . The orbits $[\mathbf{X}]$ are in 1-1-correspondence with the partitions $X = \pi(\mathbf{X})$. Therefore, we can identify partitions with orbits and \mathcal{P} with \mathcal{X}/Π . Consequently, we occasionally write $\mathbf{X} \in X$ if $X = \pi(\mathbf{X})$.

2.3. Intrinsic Metric

Next, we endow the partition space \mathcal{P} with an intrinsic metric δ related to the Euclidean distance such that (\mathcal{P}, δ) becomes a geodesic space. The Euclidean norm for matrices $\mathbf{X} \in \mathcal{X}$ is defined by

$$\|\mathbf{X}\| = \left(\sum_{k=1}^{\ell} \sum_{j=1}^m |x_{kj}|^2 \right)^{1/2}.$$

The Euclidean norm induces a distance on \mathcal{P} of the form

$$\delta : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}, \quad (X, Y) \mapsto \min \{\|\mathbf{X} - \mathbf{Y}\| : \mathbf{X} \in X, \mathbf{Y} \in Y\}.$$

Then the pair (\mathcal{P}, δ) is a geodesic metric space [13], Theorem 2.1.

3. Homogeneity of a Sample

This section first links consensus clustering to the field of Fréchet functions from Mathematical Statistics. Then we present conditions of uniqueness. Based on these conditions, we study how likely is a unique mean partition. For this, we propose homogeneity of a sample as a measure of how close a sample is to having a unique mean. We present an efficient way to compute a lower bound of homogeneity. Finally, this section relates homogeneity to cluster stability and points to potentially conflicting approaches in clustering: stability and diversity in consensus clustering.

3.1. Fréchet Functions

In this section, we link the consensus function of the mean partition approach to Fréchet functions [8]. This link provides access to many results from Statistics in Non-Euclidean spaces [1].

Let (\mathcal{P}, δ) be a partition space endowed with the metric δ induced by the Euclidean norm. We assume that Q is a probability measure on \mathcal{P} with support \mathcal{S}_Q .¹ The function

$$F_Q : \mathcal{P} \rightarrow \mathbb{R}, \quad Z \mapsto \int_{\mathcal{P}} \delta(X, Z)^2 dQ(X)$$

is the expected Fréchet function of Q . The minimum of F_Q exists but is not unique, in general [13]. Any partition $M \in \mathcal{P}$ that minimizes F_Q is an expected partition. We say Q is homogeneous, if the expected partition of Q is unique. Otherwise, Q is said to be heterogeneous. The minimum $V_Q = F_Q(M)$ is called the variation of Q .

Let $\mathcal{S}_Q^n = \mathcal{S}_Q \times \cdots \times \mathcal{S}_Q$ denote the n -fold cartesian product of support \mathcal{S}_Q . If $\mathcal{S}_n = (X_1, X_2, \dots, X_n) \in \mathcal{S}_Q^n$ is a sample of n partitions, then the (empirical) Fréchet function of \mathcal{S}_n is of the form

$$F_n : \mathcal{P} \rightarrow \mathbb{R}, \quad Z \mapsto \frac{1}{n} \sum_{i=1}^n \delta(X_i, Z)^2.$$

As for expected Fréchet functions F_Q , the minimum of F_n exists but is not unique, in general [13]. Any partition $M \in \mathcal{P}$ that minimizes F_n is a mean partition. The minimum $V_n = F_n(M)$ is the variation of \mathcal{S}_n . We say the sample \mathcal{S}_n is homogeneous, if the mean partition of \mathcal{S}_n is unique. Otherwise, \mathcal{S}_n is said to be heterogeneous.

3.2. Conditions of Uniqueness

In this section, we show that the expected and mean partition are unique if the partitions to be summarized are contained in a sufficiently small ball.

¹The support of Q is the smallest closed subset $\mathcal{S}_Q \subseteq \mathcal{P}$ such that $Q(\mathcal{S}_Q) = 1$.

The ball $\mathcal{B}(Z, r)$ with center $Z \in \mathcal{P}$ and radius r is a set of the form

$$\mathcal{B}(Z, r) = \{X \in \mathcal{P} : \delta(X, Z) \leq r\}.$$

We call the ball $\mathcal{B}(Z, r)$ homogeneous if there is a bijective isometry

$$\psi : \mathcal{B}(Z, r) \longrightarrow \mathcal{B}(\mathbf{Z}, r),$$

where $\mathcal{B}(\mathbf{Z}, r)$ is the ball in the Euclidean space \mathcal{X} centered at representation $\mathbf{Z} \in Z$. The definition of homogeneous ball is independent of the choice of representation, because two balls in \mathcal{X} at different centers but identical radius r are isometric.

The maximum homogeneity (max-hom) radius ρ_Z at Z is the largest radius for which $\mathcal{B}_Z = \mathcal{B}(Z, \rho_Z)$ is a homogeneous ball. We call \mathcal{B}_Z the max-hom ball centered at Z . The next result guarantees uniqueness of the expectation and mean if the partitions to be summarized are contained in an open subset of some max-hom ball.

Theorem 3.1. *Let Q be a probability measure on \mathcal{P} with support \mathcal{S}_Q . Suppose that there is a partition $Z \in \mathcal{P}$ and an open subset $\mathcal{U} \subset \mathcal{B}_Z$ such that $\mathcal{S}_Q \subseteq \mathcal{U}$. Then Q and any sample $\mathcal{S}_n \in \mathcal{S}_Q^n$ are homogeneous.*

Note that Theorem 3.1 makes no statement about the existence and size of max-hom balls. Therefore, it is unclear whether the uniqueness conditions are satisfied only in exceptional cases or are of practical relevance. The following treatment is devoted to this issue.

3.3. Asymmetric Partitions

This section sets the stage for understanding how likely and how feasible are unique expectations and mean partitions. To this end, we introduce the notion of asymmetric partition. Based on the notion of asymmetry, we characterize partitions with positive max-hom radius.

Let $\Pi^* = \Pi \setminus \{\mathbf{I}\}$ denote the subset of $(\ell \times \ell)$ -permutation matrices without identity matrix \mathbf{I} . The degree of asymmetry of a partition $Z \in \mathcal{P}$ is defined by

$$\alpha_Z = \min \{\|\mathbf{Z} - \mathbf{P}\mathbf{Z}\| : \mathbf{Z} \in Z \text{ and } \mathbf{P} \in \Pi^*\}.$$

A partition Z is asymmetric if $\alpha_Z > 0$. If $\alpha_Z = 0$, the partition Z is called symmetric. The next result establishes a relationship between the degree of asymmetry and the max-hom radius.

Proposition 3.2. *Let $Z \in \mathcal{P}$ be a partition. Then we have*

1. $\alpha_Z/4 \leq \rho_Z$
2. $\alpha_Z > 0 \Leftrightarrow \rho_Z > 0$.

Proposition 3.2(1) says that $\mathcal{A}_Z = \mathcal{B}(Z, \alpha/4)$ is a homogeneous ball. We call \mathcal{A}_Z the asymmetry ball of Z . From Theorem 3.1 and $\mathcal{A}_Z \subseteq \mathcal{B}_Z$ follows that expectation and mean are unique if the support \mathcal{S}_Q is contained in an open subset of \mathcal{A}_Z .

Proposition 3.2(2) states that being an asymmetric partition and having a positive max-hom radius are equivalent properties. Thus, we can characterize partitions with positive max-hom radius by asymmetric partitions:

Proposition 3.3.

1. *Almost all partitions are asymmetric.*
2. *A partition is asymmetric if and only if its clusters are mutually distinct.*

The first assertion of Prop. 3.3 states that partitions with degenerated max-hom ball that collapse to a single point are the pathological cases in the sense that they are contained in some subset of measure zero.

The second assertion of Prop. 3.3 provides us a way of how to compute the degree of asymmetry as we will see shortly. Recall that a cluster of a partition Z is represented by a row \mathbf{z}_k of a representation $\mathbf{Z} \in Z$. Empty clusters are represented by zero rows. A pair of clusters of Z is distinct if the corresponding rows of \mathbf{Z} are distinct. The next results are an immediate consequence of Prop. 3.3(2).

Corollary 3.4.

1. *Every hard partition with at most one empty cluster is asymmetric.*
2. *A symmetric partition has at least one pair of identical clusters.*
3. *A partition with more than one empty cluster is symmetric.*

Next, we show how the degree of asymmetry of a partition can be determined.

Proposition 3.5. *Let $Z \in \mathcal{P}_{m,\ell}$ be a partition.*

1. *Let $\mathbf{Z} \in Z$ be a representation with rows $\mathbf{z}_1, \dots, \mathbf{z}_\ell$. Then*

$$\alpha_Z = \min \left\{ \sqrt{2} \cdot \|\mathbf{z}_p - \mathbf{z}_q\| : 1 \leq p < q \leq \ell \right\}$$

2. *Suppose that Z is a hard partition. Then*

$$\alpha_Z = \sqrt{2(m_1 + m_2)},$$

where $m_1 \leq m_2$ are the sizes of the two smallest clusters of Z .

3. *Suppose that Z is an asymmetric hard partition. Then*

$$\sqrt{2} \leq \alpha_Z \leq 2 \cdot \sqrt{\left\lceil \frac{m}{\ell} \right\rceil}.$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x .

The first statement of Prop. 3.3 tells us how to compute the degree of asymmetry of an arbitrary partition. The second statement of Prop. 3.3 gives us a simpler formula for computing the degree of asymmetry for the subset of hard partitions. Finally, the last statement of Prop. 3.3 tells us the range of values the degree of asymmetry can take for the subset of asymmetric hard partitions. Hard partitions have largest degree of asymmetry if the data points are evenly distributed across all clusters. Conversely, the degree of asymmetry is small if there are clusters with few data points.

3.4. Homogeneity

In this section, we introduce homogeneity as a measure of how close a sample is to having a unique mean and provide a lower bound that can be easily determined.

Suppose that $\mathcal{S}_n = (X_1, \dots, X_n) \in \mathcal{S}_Q^n$ is a sample of n partitions. By $\mathcal{H}(\mathcal{S}_n)$ we denote the set of all homogeneous sub-samples of \mathcal{S}_n , that is the set of all sub-samples of \mathcal{S}_n with unique mean partition. Obviously, $\mathcal{H}(\mathcal{S}_n)$ is non-empty, because sub-samples consisting of a singleton are homogeneous. If \mathcal{S}_n is homogeneous, then $\mathcal{H}(\mathcal{S}_n)$ coincides with the power set of \mathcal{S}_n .

The homogeneity of a sample \mathcal{S}_n is defined by

$$H(\mathcal{S}_n) = \max \left\{ \frac{|S|}{n} : S \in \mathcal{H}(\mathcal{S}_n) \right\}.$$

Homogeneity measures how close a sample is to being homogeneous. Homogeneity quantifies the largest fraction of partitions that have a unique mean partition. Conversely, the value $1 - H$ tells us how many partitions we need to remove from \mathcal{S}_n to obtain a sub-sample with unique mean partition. Homogeneous samples have homogeneity one and heterogeneous samples have homogeneity less than one. In the worst case, the homogeneity of a sample \mathcal{S}_n is $H(\mathcal{S}_n) = 1/n$.

It is unclear how to compute the homogeneity $H(\mathcal{S}_n)$ efficiently. We therefore present a procedure to determine a lower bound of $H(\mathcal{S}_n)$ by using the degree of asymmetry. Let \mathcal{A}_i be the asymmetry ball of the i -th sample partition X_i . By

$$\mathbb{I}_{\mathcal{A}_i}(X_j) = \begin{cases} 1 & : X_j \in \mathcal{A}_i \\ 0 & : \text{otherwise,} \end{cases}$$

we denote the indicator function of \mathcal{A}_i . We can evaluate $\mathbb{I}_{\mathcal{A}_i}(X_j)$ by first determining the degree of asymmetry α_i of X_i according to Prop. 3.5. Then we test membership of X_j in \mathcal{A}_i by evaluating the expression

$$\delta(X_i, X_j) \leq \frac{1}{4}\alpha_i.$$

The fraction of sample partitions of \mathcal{S}_n that are contained in \mathcal{A}_i is given by

$$h_i = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_{\mathcal{A}_i}(X_j).$$

Then the approximated homogeneity (α -homogeneity) of sample \mathcal{S}_n is defined as

$$h^*(\mathcal{S}_n) = \max_i h_i.$$

Obviously, the α -homogeneity $h^*(\mathcal{S}_n)$ is a lower bound of $H(\mathcal{S}_n)$. Note that $h^*(\mathcal{S}_n) = 1$ implies that the sample \mathcal{S}_n is homogeneous and therefore has a unique mean partition. If $h^*(\mathcal{S}_n) < 1$ no statement can be made about whether \mathcal{S}_n is homogeneous or heterogeneous. In this case, $h^*(\mathcal{S}_n)$ measures how likely a unique mean is.

3.5. Clustering Stability

This section links clustering instability to consensus clustering and sketches how homogeneity is related to clustering stability in a simplified setting. Finally, we briefly point to a potential conflict between cluster stability and diversity in consensus clustering.

3.5.1. Clustering Instability

Choosing the number ℓ of clusters is a persisting model selection problem in clustering. One way to select ℓ is based on the concept of clustering stability. The intuitive idea behind clustering stability is that a clustering algorithm should produce similar partitions if repeatedly applied to slightly different datasets from the same underlying distribution.

Here, we assume that $\mathcal{S}_{n,k} = (X_1, \dots, X_n)$ is a sample of n partitions $X_i \in \mathcal{P}_{k,m}$ of (possibly different) datasets of size m with k clusters. Following [16], model selection in clustering is posed as the problem of minimizing the function

$$I_{n,k} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Delta_k(X_i, X_j)$$

over all numbers k of clusters such that $1 \leq k_{\min} \leq k \leq k_{\max} \leq m$. Then one option to choose the number ℓ of clusters is as follows:

$$\ell = \arg \min_k I_{n,k}.$$

The function $I_{n,k}$ is called cluster instability and measures the average distance between partitions. Another less common interpretation is that cluster instability measures the average variation $F_{n,k}(X_i)$ of the sample partitions X_i of $\mathcal{S}_{n,k}$, where

$$F_{n,k}(X_i) = \frac{1}{n} \sum_{j=1}^n \Delta_k(X_i, X_j)$$

is the Fréchet function of $\mathcal{S}_{n,k}$ with respect to the distance Δ_k . Thus, we can equivalently rewrite cluster instability as

$$I_{n,k} = \frac{1}{n} \sum_{i=1}^n F_{n,k}(X_i).$$

The last equation links cluster stability to consensus clustering and to Fréchet functions.

3.5.2. Homogeneity vs. Stability

Intuitively, we expect that the average pairwise distance $I_{k,m}$ between partitions and the average distance $F_{n,k}(M_k)$ to a mean partition are correlated if the underlying distance function Δ_k is well-behaved. If Δ_k is a metric, we have (see Section B.5)

$$F_{n,k}(M_k) \leq I_{n,k}, \quad (1)$$

where M_k is a mean or medoid partition of sample $\mathcal{S}_{n,k}$.² These considerations suggest that the variation $F_{n,k}(M_k)$ can serve as an alternative score function for model selection that is related to cluster instability $I_{n,k}$. We choose the number ℓ of clusters according to the rule

$$\ell = \arg \min_k F_{n,k}(M_k). \quad (2)$$

To relate homogeneity to cluster stability consider the (non-symmetric) distance

$$\Delta_k(X_i, X_j) = 1 - \mathbb{I}_{\mathcal{A}_i}(X_j).$$

The Fréchet function of $\mathcal{S}_{n,k}$ takes the form

$$F_{n,k}(Z) = \frac{1}{n} \sum_{i=1}^n 1 - \mathbb{I}_{\mathcal{A}_Z}(X_i) = 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\mathcal{A}_Z}(X_i) = 1 - h_Z,$$

where h_Z is the fraction of sample partitions of $\mathcal{S}_{n,k}$ that are contained in the asymmetry ball \mathcal{A}_Z of Z . Then we can rewrite the Fréchet variation $F_{n,k}(M_k)$ at a medoid M_k by

$$F_{n,k}(M_k) = 1 - h_k^*,$$

where h_k^* is the α -homogeneity of $\mathcal{S}_{n,k}$. Thus, choosing the number ℓ of clusters according to Equation (2) is equivalent to choosing ℓ according to

$$\ell = \arg \max_k h_k^*.$$

We choose ℓ in such a way that uniqueness of the mean partition is most likely. This shows the relationship between uniqueness of the mean partition and cluster stability.

In contrast to cluster instability, α -homogeneity measures stability with respect to the size of smallest clusters. To see this, recall that the degree of asymmetry of a partition Z is

$$\alpha_Z = \sqrt{2(m_1 + m_2)},$$

where m_1 and m_2 are the sizes of the two smallest clusters (see Prop. 3.5(2)). Then a hard partition X is in the asymmetry ball \mathcal{A}_Z of hard partition Z if both partitions disagree on at most $(m_1 + m_2)/4$ data points. This shows that a clustering is as stable as the smallest clusters in its partition.

²A medoid is a sample partition $M_k \in \mathcal{S}_{k,m}$ such that $F_{n,k}(M_k) \leq F_{n,k}(X_i)$ for all $1 \leq i \leq n$.

3.5.3. Stability vs. Diversity

In consensus clustering it is recommended to use diverse partitions to improve the performance [5, 22, 23]. Following [5], diversity is measured in the same way as cluster instability, namely by the sum of pairwise distances between sample partitions.

Though diversity corresponds to cluster instability, its application in consensus clustering does not contradict the goal of cluster stability per se. What matters – from the point of view of cluster stability – is whether the resulting consensus partitions are stable. Since diversity corresponds to low homogeneity, it is unlikely that a sample of diverse partitions has a unique mean. To comply with cluster stability, the question is under which conditions are two different mean partitions similar? To answer this question, we need the following result proved by [3]:

Theorem 3.6. *Let $M \in \mathcal{P}$ is a mean partition of the sample $\mathcal{S}_n = (X_1, \dots, X_n) \in \mathcal{P}^n$. Then every representation $\mathbf{M} \in M$ is of the form*

$$\mathbf{M} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

where $\mathbf{X}_i \in X_i$ are in optimal position with \mathbf{M} , that is $\delta(X_i, M) = \|\mathbf{X}_i - \mathbf{M}\|$.

Theorem 3.6 describes the form of a mean partition in terms of representations of the sample partition. Since every partition has only finitely many different representations, the set of mean partitions of a given sample is finite and therefore discrete.

Now suppose that the sample \mathcal{S}_n has two different mean partitions M and M' . Let $\mathbf{M} \in M$ and $\mathbf{M}' \in M'$ be representations of both mean partitions in optimal position. Applying Theorem 3.6 gives

$$\delta(M, M') = \|\mathbf{M} - \mathbf{M}'\| = \left\| \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i - \frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \right\| = \frac{1}{n} \left\| \sum_{i=1}^n (\mathbf{X}_i - \mathbf{X}'_i) \right\|$$

where $\mathbf{X}_i, \mathbf{X}'_i \in X_i$ are representations in optimal position with \mathbf{M} and \mathbf{M}' , respectively. Since both means are different, there is a non-empty subset $\mathcal{J} \subseteq \{1, \dots, n\}$ of indices such that $\mathbf{X}_j \neq \mathbf{X}'_j$ for all $j \in \mathcal{J}$. We obtain

$$\delta(M, M') \leq \frac{1}{n} \sum_{i=1}^n \|\mathbf{X}_i - \mathbf{X}'_i\| = \frac{1}{n} \sum_{j \in \mathcal{J}} \|\mathbf{X}_j - \mathbf{X}'_j\|.$$

Consider the following conditions:

1. The index set \mathcal{J} is small, that is $|\mathcal{J}| \ll n$,
2. The degree of asymmetry α_j of partitions X_j is low for all $j \in \mathcal{J}$,
3. The distance $\|\mathbf{X}_j - \mathbf{X}'_j\|$ is close to α_j for all $j \in \mathcal{J}$.

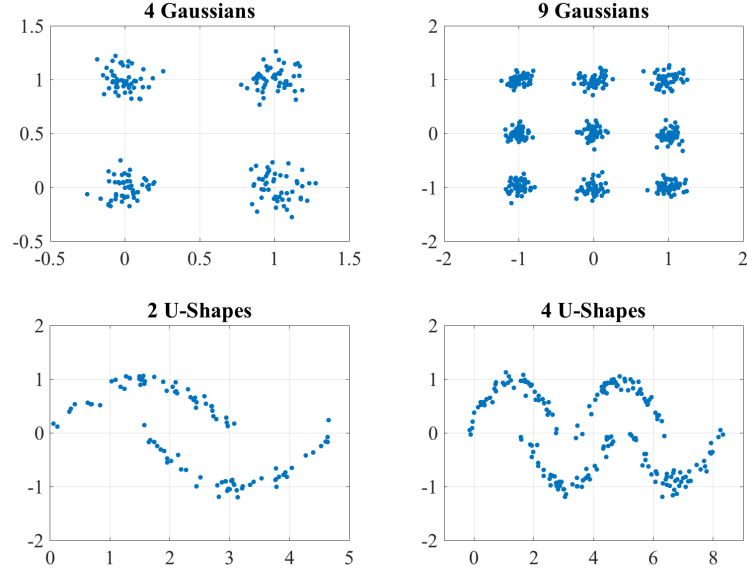


Figure 1: Examples of the G4, G9, U2, and U4 dataset.

Then two different mean partitions are similar if condition (1) or if both conditions (2) and (3) hold. Under these conditions, cluster stability and diversity in consensus clustering are not conflicting approaches.

In general, it is not self-evident that different mean partitions of a sample of diverse partitions are similar. Therefore, we point to the possibility that cluster stability and diversity in consensus clustering can be contradictory approaches in achieving the common goal of improved cluster performance.

4. Experiments

The goal of this section is to assess the homogeneity of samples obtained by k-means applied to synthetic and real-world data.

4.1. Experiments on Synthetic Data

Data. We generated the following types of datasets in \mathbb{R}^2 :

1. UD: Uniform distribution
2. G4: Four Gaussians
3. G9: Nine Gaussians

4. U2: Two U-Shapes

5. U4: Four U-Shapes

The UD datasets consists of m data points drawn from the uniform distribution on the unit square. The G4 and G9 dataset consists of m data points drawn from four and nine Gaussian distributions, resp., with identical covariance matrix $\sigma^2 \mathbf{I}$ and different mean vectors. The mean vectors of the G2 dataset are the four vertices of the unit square. The nine mean vectors of the G9 dataset are of the form (x, y) with $x, y \in \{-1, 0, 1\}$. The U2 and U4 dataset consists of m data points forming concave and convex U-Shapes. The data points were generated by imposing Gaussian noise with mean zero and standard deviation σ on the positive and negative component of the sine-function. The U2 dataset has one concave and one convex U-Shape, whereas the U4 dataset has two of both types of U-Shapes.

Data points were evenly distributed across the different clusters of the G4, G9, U2, and U4 datasets. Figure 1 shows examples of all datasets with the exception of the UD dataset. The parameters for all datasets were $\sigma = 0.1$ and $m_c = 50$, where m_c is the number of data points of a single cluster (with $m_c = 50$, we have $m_{UD} = 50$, $m_{G4} = 200$, $m_{G9} = 450$, $m_{U2} = 100$, and $m_{U4} = 200$).

Generic Protocol. A single experiment was conducted according to the following generic scheme:

Input:

- m – number of data points
- σ – standard deviation
- k – parameter of k-means

Procedure:

- Generate a dataset \mathcal{Z} of size m with standard deviation σ
- Repeat $n = 100$ times:
 - Apply the k-means algorithm to dataset \mathcal{Z} to obtain sample $\mathcal{S}_{n,k}$
 - Compute the α -homogeneity $h^*(\mathcal{S}_{n,k})$

Output:

- α -homogeneity $h^*(\mathcal{S}_{n,k})$
-

The procedure was repeated 100-times using the same input parameters. Finally, the average α -homogeneity h^* over the 100 trials was recorded.

The UD datasets served as a base-line. For these datasets, σ is a factor with which the uniformly generated data points were multiplied.

4.1.1. Results on G4 Datasets

The goal of the first series of experiments is to assess homogeneity as a function of the parameters k , σ , and m under the assumption that the cluster structure in the data can be essentially discovered by the k-means algorithm for a suitable value of k .

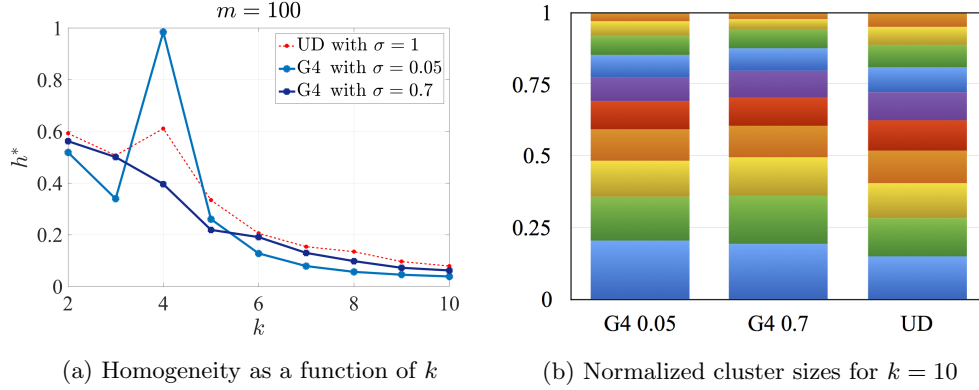


Figure 2: Average α -homogeneities h^* as a function of the k-means parameter k and normalized cluster sizes for $k = 10$ ordered from largest (bottom) to smallest (top), respectively.

We considered G4 datasets and contrasted the results to those obtained on UD datasets. Unless otherwise stated, the default input parameters were

- $k = 4$ for the k-means algorithm,
- $m = 100$ for the size of the datasets,
- $\sigma = 1$ as factor for the UD datasets.

Homogeneity as a function of k . We considered three types of datasets: (i) G4 generated with standard deviation $\sigma = 0.05$, (ii) G4 generated with standard deviation $\sigma = 0.7$, and (iii) UD generated with factor $\sigma = 1$. For every $k \in \{2, \dots, 10\}$ and for all three types of datasets, we conducted experiments according to the above described generic protocol.

Figure 2a shows the average α -homogeneities h^* as a function of the number k . We made the following observations:

1. The general trend is that homogeneity decreases with increasing k . To understand why homogeneity decreases with increasing k , recall that the degree of asymmetry of a partition Z is

$$\alpha_Z = \sqrt{2(m_1 + m_2)},$$

where m_1 and m_2 are the sizes of the two smallest clusters (see Prop. 3.5(2)). From the strong form of the pigeonhole principle follows

$$\frac{m_1 + m_2}{2} \leq \frac{m}{k}.$$

Thus, the sum $m_1 + m_2$ decreases with increasing number k of clusters. This means that homogeneity is likely to be lower for large k given a fixed number m of data

points. Imbalanced cluster sizes further deteriorate the situation. Figure 2b shows the average cluster sizes of k-means with $k = 10$ for all three types of datasets. The cluster sizes are normalized by m . We see that the cluster sizes of all types of datasets are imbalanced, in particular both types of the G4 datasets. These findings indicate that increasing the parameter k results in increasingly less stable clusterings and makes a unique mean partition increasingly less likely.

2. The trend is interrupted at $k = 4$ for UD datasets and for G4 datasets with $\sigma = 0.05$. For $\sigma = 0.05$ the G4 dataset has a clearly visible cluster structure (see Figure 3). The k-means algorithm with $k = 4$ recovers this structure resulting in high homogeneity. Only a few partitions need to be removed in order to guarantee uniqueness of the mean. For $\sigma = 0.7$ no cluster structure is visible as shown in Figure 3. Consequently, nothing unexpected happened and homogeneity is low. These findings suggest that uniqueness of the mean partition is more likely when k-means is capable to essentially discover the cluster structure of the dataset.

Surprisingly, there is a moderate peak at $k = 4$ on UD datasets conveying that k-means is most stable when viewing uniformly distributed data as a 2, and 4 cluster problem (homogeneity of samples of partitions with one cluster is always one). This moderate peak is even more notable when compared to the results on G4 datasets with $\sigma = 0.7$. This result indicates that peaks in homogeneity do not necessarily allow us to draw conclusions about the cluster structure in a dataset.

By combining both findings, we hypothesize that an evident cluster structure discovered by the underlying algorithm implies peaks in homogeneity but the converse claim does not necessarily hold.

3. For $k \neq 4$ homogeneity is lower the more structure we assume in the dataset. We assume higher structure in G4 datasets with lower variance and we assume higher structure in any G4 datasets than in UD datasets. The results show that the more evident the assumed cluster structure is the lower is the homogeneity for a wrong choice of k . These findings suggest that mismatching an evident cluster structure in the data can introduce additional instability into the clusterings. Further research is necessary to test this hypothesis.

Homogeneity as a function of σ . For every $\sigma \in \{0.01, 0.05, 0.1, 0.2, \dots, 1.0\}$ and for both datasets of type G4 and UD, we conducted experiments according to the above described generic protocol. Figure 3 depicts examples of G4 datasets with varying standard deviation σ .

Figure 4 shows the average α -homogeneity h^* as a function of the standard deviation σ . We observed that the average α -homogeneity on G4 datasets decreases with increasing standard deviation σ until saturation. Homogeneity is at a high level with values above 0.9, when the four clusters are clearly visible. For larger standard deviations, the clusters become increasingly blurred and the average α -homogeneity rapidly drops below 0.4. The turning point between high and low homogeneity is around $\sigma = 0.3$ and roughly corresponds to the subjective turning point of what we might perceive as a visible cluster structure (cf. Figure 3). Moreover, homogeneity on G4 datasets around

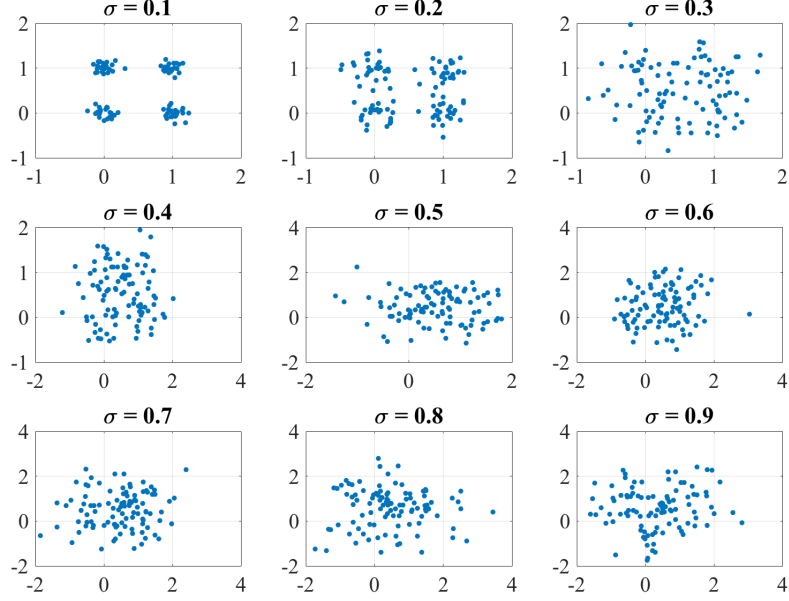


Figure 3: Each plot shows 100 data points randomly generated by four Gaussian distributions with identical mean and increasing standard deviation σ .

the turning point is comparable to homogeneity on UD datasets, which is invariant under scaling of σ . Based on these results we raise the hypothesis that an algorithm that essentially discovers a visible cluster structure in the data is stable and guarantees a unique mean partition after removing a small fraction of partitions.

Homogeneity as a function of m . We considered three types of datasets: (i) G4 with $\sigma = 0.3$, (ii) G4 with $\sigma = 0.7$, and (iii) UD with factor $\sigma = 1$. Let m_c denote the number of data points of component c of a dataset. Then G4 datasets have size $m = 4m_c$ and UD datasets are of size $m = m_c$. For every

$$m_c \in \{25, 50, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 5000\}$$

and for all three types of datasets, we conducted experiments according to the above described generic protocol.

Figure 5 shows an excerpt of the average α -homogeneities h^* as a function of the number m_c of data points in each component. We observed that homogeneity increases with increasing dataset size m . Moreover, homogeneity increases faster and more substantially for G4 datasets with lower standard deviation. If the four clusters are just visible as for $\sigma = 0.3$, the mean partition is likely to be unique for datasets with

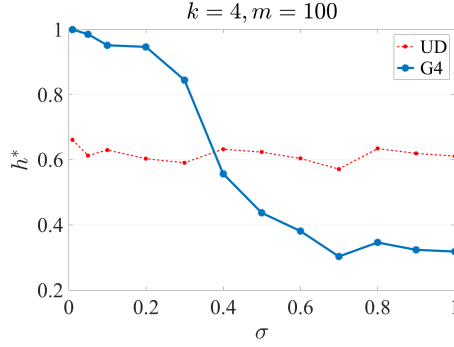


Figure 4: Average α -homogeneity as a function of σ .

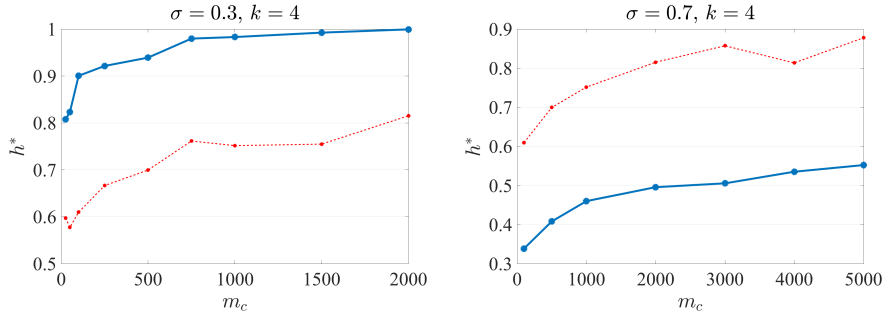


Figure 5: Average α -homogeneity as a function of the number m_c of data points per cluster. The blue (red) line refers to results on the G4 (UD) datasets.

at least $m_c = 2000$ data points in each cluster. Surprisingly, homogeneity is high even for UD datasets with more than 5,000 data points. These results indicate that homogeneity can be improved and uniqueness of the mean partition can be eventually enforced by increasing the size of the dataset.

4.1.2. Results on U-Shapes and Gaussians

The goal of the second series of experiments is to assess homogeneity as a function of the parameter k under the assumption that k-means is unable to essentially discover a visible cluster structure in the data. For this we considered U2 and U4 datasets and contrasted the results to those obtained on G4 and G9 datasets. The parameters for all datasets were $\sigma = 0.1$ and $m_c = 50$, where m_c is the number of data points of a single cluster. Figure 1 shows examples of all four datasets.

Figure 6 shows the average α -homogeneities for all four datasets. As before, the general trend is that homogeneity decreases with increasing k and is only interrupted when the clearly visible cluster structure in the dataset can be discovered by the k-

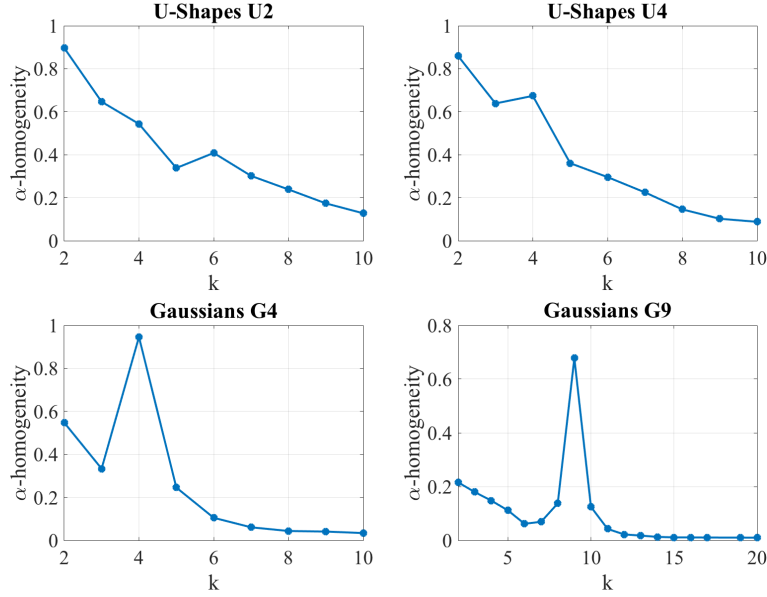


Figure 6: Average α -homogeneity as a function of the parameter k .

means algorithm. The peaks at $k = 4$ and $k = 9$ are evident for the G4 and G9 datasets, respectively. This shows that k -means is most stable and a unique mean partition is most likely when the parameter k coincides with the number of clusters in the datasets.

The situation is different for both U-Shape datasets. The k -means algorithm is not able to essentially discover the cluster structure of the U2 and U4 datasets. Nevertheless, homogeneity is at the highest level for $k = 2$ for both datasets. While this result is desirable for the U2 dataset at the first glance, it is unsatisfactory for the U4 dataset. The result for the U2 dataset with $k = 2$ is only desirable at the first glance for the following reason: Although k -means performed stable and only a small fraction of partitions need to be removed in order to guarantee a unique mean partition, the discovered cluster structure does not properly match with the underlying cluster structure in the data as indicated by the top-left plot of Figure 7. The same holds for the U4 dataset as shown by the top-right plot of Figure 7. Though the result for the U4 dataset with $k = 2$ is unsatisfactory, closer inspection of the plots reveals that homogeneity has a moderate peak at $k = 4$ for the U4 dataset not present for the U2 dataset. A further small peak is at $k = 6$ for the U2 dataset. In the latter case, k -means frequently refines each of the two U-Shapes into three clusters (c.f. plots at bottom row of Figure 7). Such peaks may indicate that there could be a cluster structure but the underlying algorithm is unable to essentially discover this structure. A counter-example for this claim is the peak at $k = 4$ for UD datasets as discussed in the

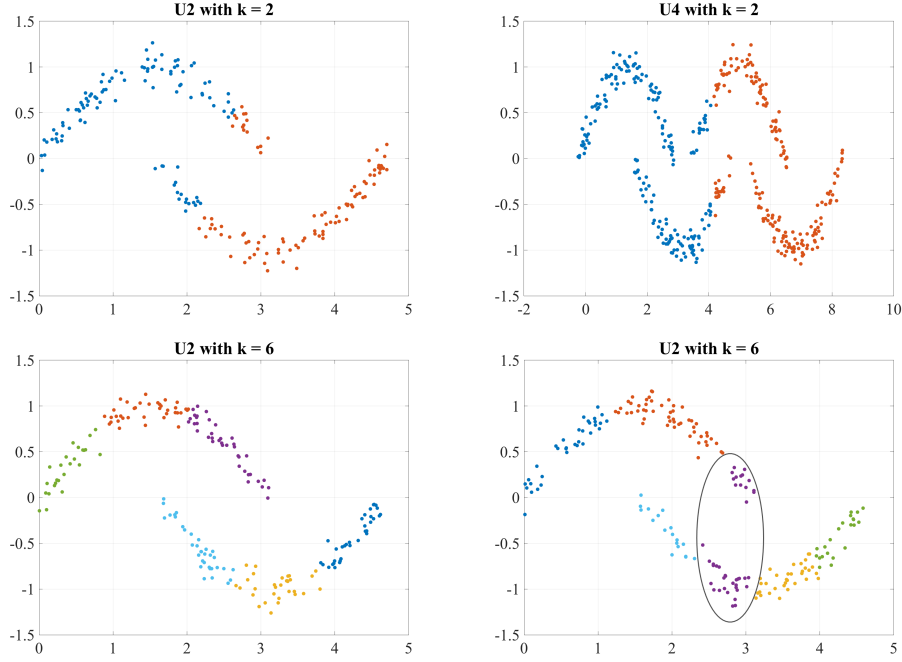


Figure 7: Examples of partitions obtained by the k-means algorithm for the U2 and U4 dataset. Top row shows typical results for k-means with $k = 2$ on both datasets. Bottom row shows results of k-means with $k = 6$ on U2 datasets. The bottom-left plot shows a partition that is a refinement of the U2 cluster structure, which is not the case for the bottom-right plot due to the cluster circumscribed by the ellipse.

previous experiment. The conclusion is that high homogeneity and stability are merely indicators for a possible cluster structure in the data but need further examination.

4.2. Experiments on UCI Datasets

The goal of this experiment is to investigate how likely is a unique mean partition for real-world datasets. For this, we considered six datasets from the UCI Machine Learning Repository [15] listed in Table 1. For every dataset and for every $k \in \{2, \dots, 10\}$, we applied k-means 100-times and recorded the α -homogeneity.

Figure 8 shows the α -homogeneities for each dataset as a function of the number k . We observed that (i) uniqueness of the mean partition is guaranteed for small values of k , and (ii) homogeneity decreases with increasing k . Exceptions from these general observations are the music and eye dataset.

Except for the music dataset, observation (i) shows that uniqueness of the mean partition can be guaranteed for real world data sets. This result indicates that uniqueness

Dataset	Abbr.	Classes	Elements	Features
Banknote Authentication	bank	2	1372	4
EEG Eye State	eye	2	14980	14
Iris	iris	3	150	4
Geographical Original of Music	music	33	1059	68
Pima Indians Diabetes	pima	2	768	8
Connectionist Bench	sonar	2	208	60

Table 1: Characteristics of six datasets from the UCI Machine Learning Repository.

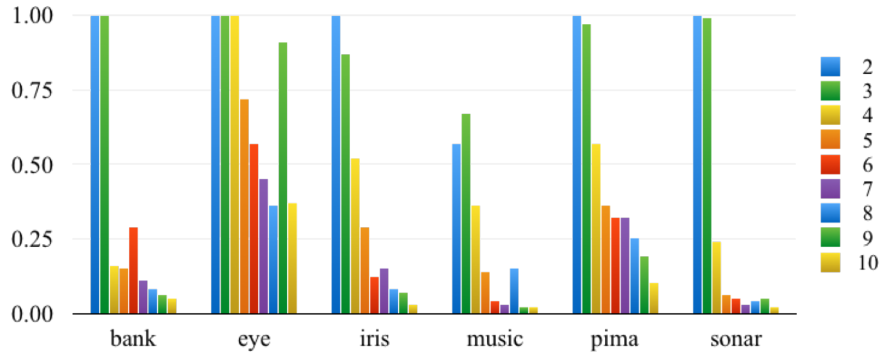


Figure 8: α -Homogeneities for each dataset as a function of the number k .

is of practical relevance and not a matter of exceptional cases.

Observation (ii) is in line with the results on synthetic data and an explanation follows the same argumentation as in Section 4.1.1. As shown in Figure 9, the clusters are highly unbalanced for $k = 10$. Consequently, uniqueness of the mean partition is less likely for large k . Increasing the number m of data points can improve homogeneity. This is possibly one reason why the eye dataset has substantially larger α -homogeneity than the other datasets for all k and has an additional peak of high homogeneity for $k = 9$, although it has the most imbalanced partitions for $k = 10$.

5. Conclusion

Uniqueness of the mean partition is a desirable property in consensus clustering that comes along with several benefits. We showed that both, the expected partition and the mean partition, are unique when the support is contained in an open subset of some max-hom ball. According to this condition, uniqueness is neither an exceptional nor a generic property. To cope with this issue, we proposed homogeneity as a measure of how close a sample is to having a unique mean. Homogeneity is not confined to consensus clustering but is also related to cluster stability. This in turn points to the possibility that cluster stability and diversity in consensus clustering can be conflicting

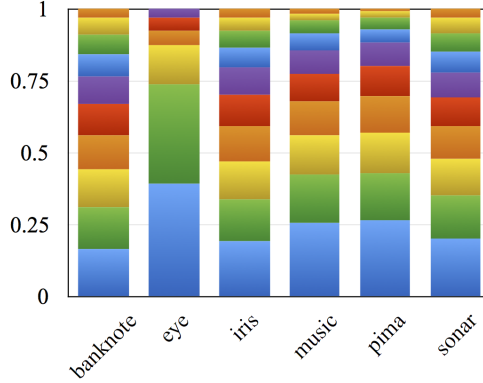


Figure 9: Normalized cluster sizes for $k = 10$.

goals. Homogeneity can be efficiently bounded from below by α -homogeneity, which applies the degree of asymmetry of a partition. With α -homogeneity, we can identify a sub-sample of the largest sub-sample of partitions that can be retained in order to guarantee a unique mean partition. Preliminary empirical results show that uniqueness occurs in real-world data and can be enforced by increasing the size of the dataset or by removing outlier partitions when α -homogeneity is high. The results also indicate that α -homogeneity can be used as a criterion for model selection. The results of this paper can be placed into the general context of a statistical theory of partitions, which embraces consensus clustering and cluster stability as special cases.

The main limitations of the proposed approach are twofold: restriction to the intrinsic metric on partitions derived from the Euclidean distance and restriction of uniqueness conditions on max-hom balls. Generalizing both restrictions are two possible directions of future research. Another important issue is to understand when cluster stability and diversity in consensus clustering collide.

A. Preliminaries

This section presents technicalities useful for proving the results proposed in the main text.

A.1. Notations

We use the following notations: By $\overline{\mathcal{U}}$ we denote the closure of a subset $\mathcal{U} \subseteq \mathcal{X}$, by $\partial\mathcal{U}$ the boundary of \mathcal{U} , and by \mathcal{U}° the open subset $\overline{\mathcal{U}} \setminus \partial\mathcal{U}$. The action of permutation $\mathbf{P} \in \Pi$ on the subset $\mathcal{U} \subseteq \mathcal{X}$ is the set defined by $\mathbf{P}\mathcal{U} = \{\mathbf{P}\mathbf{X} : \mathbf{X} \in \mathcal{U}\}$. A transposition is a permutation matrix $\mathbf{P} \in \Pi$ that permutes exactly two rows. A basic result from algebra is that any permutation matrix $\mathbf{P} = \Pi^*$ can be written as a matrix product $\mathbf{P} = \mathbf{Q}_1 \cdots \mathbf{Q}_t$ of transpositions $\mathbf{Q}_i \in \Pi$ with minimum number $t > 0$ of factors.

A.2. Dirichlet Fundamental Domains

A subset \mathcal{F} of \mathcal{X} is a fundamental set for Π if and only if \mathcal{F} contains exactly one representation \mathbf{X} from each orbit $[\mathbf{X}] \in \mathcal{X}/\Pi$. A fundamental domain of Π in \mathcal{X} is a closed set $\mathcal{F} \subseteq \mathcal{X}$ that satisfies

1. $\mathcal{X} = \bigcup_{P \in \Pi} P\mathcal{F}$
2. $P\mathcal{F}^\circ \cap \mathcal{F}^\circ = \emptyset$ for all $P \in \Pi^*$.

Proposition A.1. *Let \mathbf{Z} be a representation of an asymmetric partition $Z \in \mathcal{P}$. Then*

$$\mathcal{D}_{\mathbf{Z}} = \{\mathbf{X} \in \mathcal{X} : \|\mathbf{X} - \mathbf{Z}\| \leq \|\mathbf{X} - P\mathbf{Z}\| \text{ for all } P \in \Pi\}$$

is a fundamental domain, called Dirichlet fundamental domain centered at \mathbf{Z} .

Proof. [17], Theorem 6.6.13. ■

The next result list some properties of Dirichlet fundamental domains.

Proposition A.2. *Let $\mathcal{D}_{\mathbf{Z}}$ be a Dirichlet fundamental domain centered at representation \mathbf{Z} of an asymmetric partition $Z \in \mathcal{P}$. Then the following properties hold:*

1. *There is a fundamental set $\mathcal{F}_{\mathbf{Z}}$ such that $\mathcal{D}_{\mathbf{Z}}^\circ \subseteq \mathcal{F}_{\mathbf{Z}} \subseteq \mathcal{D}_{\mathbf{Z}}$.*
2. *We have $\mathbf{Z} \in \mathcal{D}_{\mathbf{Z}}^\circ$.*
3. *Every point $\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}^\circ$ represents an asymmetric partition.*
4. *Suppose that $\mathbf{X}, P\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}$ for some $P \in \Pi^*$. Then $\mathbf{X}, P\mathbf{X} \in \partial\mathcal{D}_{\mathbf{Z}}$.*
5. *$P\mathcal{D}_{\mathbf{Z}} = \mathcal{D}_{P\mathbf{Z}}$ for all $P \in \Pi$.*

Proof. The proof follows [12], Prop. 3.13 but is adapted to the notation and terminology of this contribution.

1. [17], Theorem 6.6.11.

2. Since Z is asymmetric, we have $\|\mathbf{Z} - \mathbf{Z}\| < \|\mathbf{Z} - P\mathbf{Z}\|$ for all $P \in \Pi^*$. This shows that $\mathbf{Z} \in \mathcal{D}_{\mathbf{Z}}^\circ$.

3. Let $\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}^\circ$ be a representation of partition X . Suppose that X is symmetric. Then there is a $P \in \Pi^*$ with $\mathbf{X} = P\mathbf{X}$. This implies $\mathbf{X} \in P\mathcal{D}_{\mathbf{Z}} \cap \mathcal{D}_{\mathbf{Z}}$. Then $\mathbf{X} \in \partial\mathcal{D}_{\mathbf{Z}}$ is a boundary point of $\mathcal{D}_{\mathbf{Z}}$ by [17], Theorem 6.6.4. This contradicts our assumption that $\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}^\circ$ and shows that \mathbf{X} is asymmetric.

4. From $\mathbf{X}, P\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}$ follows $\|\mathbf{X} - \mathbf{Z}\| = \|P\mathbf{X} - \mathbf{Z}\|$. Since Π acts by isometries, we have $\|\mathbf{X} - \mathbf{Z}\| = \|P\mathbf{X} - P\mathbf{Z}\|$. Thus, we have $\|P\mathbf{X} - \mathbf{Z}\| = \|P\mathbf{X} - P\mathbf{Z}\|$. This shows that $P\mathbf{X} \in \partial\mathcal{D}_{\mathbf{Z}}$. Let $P' \in \Pi$ be the inverse of P . Since $P \neq I$, we have $P' \neq I$. Then

$$\|\mathbf{X} - \mathbf{Z}\| = \|P\mathbf{X} - \mathbf{Z}\| = \|P'P\mathbf{X} - P'\mathbf{Z}\| = \|\mathbf{X} - P'\mathbf{Z}\|,$$

From $\|\mathbf{X} - \mathbf{Z}\| = \|\mathbf{X} - P'\mathbf{Z}\|$ follows $\mathbf{X} \in \partial\mathcal{D}_{\mathbf{Z}}$.

5. Let $\mathbf{X} \in P\mathcal{D}_{\mathbf{Z}}$. We have $\|\mathbf{X} - P\mathbf{Z}\| \leq \|P'\mathbf{X} - P\mathbf{Z}\|$ for all $P' \in \Pi$ showing that $\mathbf{X} \in \mathcal{D}_{P\mathbf{Z}}$. Now assume that $\mathbf{X} \in \mathcal{D}_{P\mathbf{Z}}$. Let $P' \in \Pi$ be the inverse of P . Then we have

$$\|\mathbf{X} - P\mathbf{Z}\| = \|P'\mathbf{X} - P'P\mathbf{Z}\| = \|P'\mathbf{X} - \mathbf{Z}\|$$

by isometry of P' . Hence, $P'\mathbf{X} \in \mathcal{D}_{\mathbf{Z}}$ and therefore $PP'\mathbf{X} = \mathbf{X} \in P\mathcal{D}_{\mathbf{Z}}$. ■

A.3. Cross Sections

Suppose that $\mathcal{D}_{\mathbf{Z}}$ is the Dirichlet fundamental domain centered at representation \mathbf{Z} of an asymmetric partition $Z \in \mathcal{P}$. A map $\mu : \mathcal{P} \rightarrow \mathcal{D}_{\mathbf{Z}}$ is a cross section into $\mathcal{D}_{\mathbf{Z}}$, if $\pi(\mu(X)) = X$ for all partitions $X \in \mathcal{P}$.

Proposition A.3. *Let $\mu : \mathcal{P} \rightarrow \mathcal{D}_{\mathbf{Z}}$ be a cross section into a Dirichlet fundamental domain $\mathcal{D}_{\mathbf{Z}}$ centered at representation \mathbf{Z} of an asymmetric partition $Z \in \mathcal{P}$. Then the following properties hold:*

1. μ is injective.
2. $\mu(\mathcal{P})$ is a fundamental set.
3. μ is a measurable mapping.

Proof. Both assertions directly follow from the definitions of cross section and fundamental set. max-hom of μ directly follows from the property $\pi \circ \mu = \text{id}$. Again from $\pi \circ \mu = \text{id}$ it follows that μ maps partitions to representations. Finally, since μ is injective, the image $\mu(\mathcal{P})$ contains exactly one representation of each partition. Hence, $\mu(\mathcal{P})$ is a fundamental set. Finally, μ is measurable, because $\mu^{-1} = \pi$ and π is an open mapping. ■

Let $(\mathcal{P}, \mathcal{B}, Q)$ be a measurable space. A cross section $\mu : \mathcal{P} \rightarrow \mathcal{D}_{\mathbf{Z}}$ is a measurable map that gives rise to a measurable space $(\mathcal{D}_{\mathbf{Z}}, \mathcal{B}_{\mu}, q)$.

B. Proofs

B.1. Proof of Theorem 3.1

Parts 1–4 show uniqueness of the expected partition and Part 5 shows uniqueness of the mean partition.

1. Both assertions trivially hold for asymmetric partitions Z , because $\mathcal{S}_Q \subseteq \mathcal{B}_Z = \{Z\}$.
2. Let $Z \in \mathcal{P}$ be an asymmetric partition such that $\mathcal{S}_Q \subseteq \mathcal{B}_Z$. We select an arbitrary representation $\mathbf{Z} \in Z$ and an arbitrary cross section $\mu : \mathcal{P} \rightarrow \mathcal{D}_{\mathbf{Z}}$. Let $\mathcal{S}_{\mathbf{Z}} = \mu(\mathcal{S}_Q)$ be the image of the support \mathcal{S}_Q . Since \mathcal{B}_Z is a homogeneous ball, we have

$$F(Y) = \int_{\mathcal{P}} \delta(X, Y)^2 dQ(X) = \int_{\mathcal{S}_Q} \delta(X, Y)^2 dQ(X) = \int_{\mathcal{S}_{\mathbf{Z}}} \|\mu(X) - \mu(Y)\|^2 dq(\mu(X)),$$

where q is the image measure of measure Q under cross section μ . The function

$$f(\mathbf{Y}) = \int_{\mathcal{S}_{\mathbf{Z}}} \|\mathbf{X} - \mathbf{Y}\|^2 dq(\mathbf{X})$$

has a unique minimum $\mathbf{M} \in \mathcal{X}$ representing partition $M \in \mathcal{P}$. From

$$\mathcal{S}_{\mathbf{Z}} \subseteq \mathcal{B}(\mathbf{Z}, \alpha_Z/4) \subsetneq \mathcal{D}_{\mathbf{Z}}^{\circ}$$

together with Prop. A.3 follows that partition M is independent of the choice of a particular cross section along \mathbf{Z} . From Prop. A.2 follows that partition M is independent of the choice of a particular representation of Z . It remains to show that partition M of the second part of

this proof is independent of the choice of partition Z that satisfies $\mathcal{S}_Q \subseteq \mathcal{B}_Z$. This is proved in the sequel.

3. Suppose that $Z' \in \mathcal{P}$ is a partition satisfying $\mathcal{S}_Q \subseteq \mathcal{B}_{Z'}$. Let $Z' \in Z'$ be a representation such that $Z, Z' \in \mathcal{D}_Z \cap \mathcal{D}_{Z'}$ and let $\mu' : \mathcal{P} \rightarrow \mathcal{D}_{Z'}$ be a cross section. By $\mathcal{S} = \mu(\mathcal{S}_Q)$ and $\mathcal{S}' = \mu'(\mathcal{S}_Q)$ we denote the images of \mathcal{S}_Q under the cross sections μ and μ' , respectively.

We show that there is a permutation matrix $P \in \Pi$ such that $\mathcal{S}' = P\mathcal{S}$. Observe that the composition $\mu' \circ \pi : \mathcal{S} \rightarrow \mathcal{S}'$ is bijective. Let $X \in \mathcal{S}$ and $X' \in \mathcal{S}'$ be representations such that $X' = \mu'(\pi(X))$. Since μ' is a cross section, both elements X and X' represent the same partition $\pi(X) \in \mathcal{S}_Q$. Then there is a permutation matrix $P \in \Pi$ such that $X' = PX$. We assume that there are representations $Y \in \mathcal{S}$ and $Y' \in \mathcal{S}'$ such that

$$Y' = \mu'(\pi(Y)) \neq PY.$$

Since Y and Y' represent the same partition $\pi(Y) \in \mathcal{S}_Q$, there is another permutation matrix $Q \in \Pi \setminus \{P\}$ such that $Y' = QY$. We find that $PY \neq QY$. To see this, observe that $Y \in \mathcal{S} \subset \mathcal{D}_Z^\circ$ is an interior point of \mathcal{D}_Z . From Prop. A.2 follows that $Y = \pi(Y)$ is asymmetric. This implies that $Y \neq RY$ for all $R \in \Pi^*$. Since $P \neq Q$, we obtain $PY \neq QY$.

Although $PY \neq QY$, we have

$$\|X - Y\| = \|PX - PY\| = \|PX - QY\|. \quad (3)$$

The first equation holds, because Π acts isometrically on \mathcal{X} . The second equation follows from $X', Y' \in \mathcal{B}_{Z'}$ together with the fact that $\mathcal{B}_{Z'}$ is a homogeneous ball. From Equation (3) follows that the Dirichlet fundamental domain $\mathcal{D}_{X'}$ centered at $X' = PX$ contains PY and QY . Then by Prop. A.2 both representations PY and QY are elements of the boundary of $\mathcal{D}_{X'}$. By assumption, only $QY \in \mathcal{S}'$ is an element of $\mathcal{B}_{Z'}$. Moreover, since $\mathcal{S}' \subsetneq \mathcal{B}_{Z'}^\circ$, we have

$$\mathcal{V} = \mathcal{D}_{QX}^\circ \cap \mathcal{B}_{Z'} \neq \emptyset.$$

Suppose that $V \in \mathcal{V}$ is a representation of partition V . Then $\delta(X, V) < \|X' - V\|$, because V is in the interior of \mathcal{D}_{QX} and X' is in the interior of $\mathcal{D}_{X'}$. This contradicts the assumption that $\mathcal{B}_{Z'}$ is a homogeneous ball isometric to $\mathcal{B}_{Z'}$. Hence, $Y' = PY$ and therefore $\mathcal{S}' = P\mathcal{S}$.

4. This part shows that partition M is independent of the choice of Z . A permutation matrix $P \in \Pi$ gives rise to a diffeomorphism

$$L_P : \mathcal{X} \rightarrow \mathcal{X}, \quad X \mapsto P^{-1}X$$

with Jacobi matrix $L'_P = P^{-1}$ and $|\det L'_P| = |\det P^{-1}| = 1$. Let q' be the image measure of Q under the cross section μ' . From

$$\mu'(\mathcal{S}_Q) = \mathcal{S}' = P\mathcal{S} = L_P^{-1} \circ \mu(\mathcal{S}_Q)$$

follows $\mu = L_P \circ \mu'$. Then we have

$$q = \mu(Q) = L_P \circ \mu'(Q) = L_P(q'),$$

that is q is the image measure of q' under L_P . For every $Y \in \mathcal{S}$, we define the function

$$g_Y(X) = \|X - Y\|^2.$$

Then we rewrite $f(Y)$ by

$$f(Y) = \int_{\mathcal{S}} \|X - Y\|^2 dq(X) = \int_{\mathcal{S}} g_Y(X) dq(X).$$

Applying the transformation formula for integrals gives

$$f(\mathbf{Y}) = \int_{\mathcal{S}} g_{\mathbf{Y}}(\mathbf{X}) dq(\mathbf{X}) = \int_{\mathcal{S}'} g_{\mathbf{Y}} \circ L_{\mathbf{P}}(\mathbf{X}') \cdot |\det L'_{\mathbf{P}}| dq'(\mathbf{X}') = \int_{\mathcal{S}'} \|\mathbf{P}^{-1}\mathbf{X}' - \mathbf{Y}\|^2 dq'(\mathbf{X}').$$

Since Π acts isometrically on \mathcal{X} , we have

$$f(\mathbf{Y}) = \int_{\mathcal{S}'} \|\mathbf{P}\mathbf{P}^{-1}\mathbf{X}' - \mathbf{P}\mathbf{Y}\|^2 dq'(\mathbf{X}') = \int_{\mathcal{S}'} \|\mathbf{X}' - \mathbf{P}\mathbf{Y}\|^2 dq'(\mathbf{X}') = f'(\mathbf{P}\mathbf{Y}).$$

From $\mathbf{Y} \in \mathcal{S}$ and $\mathcal{S}' = \mathbf{P}\mathcal{S}$ follows $\mathbf{P}\mathbf{Y} \in \mathcal{S}'$. Then the unique minimizer \mathbf{M} of $f(\mathbf{Y})$ gives $\mathbf{P}\mathbf{M}$ as the unique minimizer of $f'(\mathbf{Y}')$ on \mathcal{S}' . Both elements \mathbf{M} and $\mathbf{P}\mathbf{M}$ represent the same partition M . This shows that M is independent of the choice of Z such that $\mathcal{S}_Q \subseteq \mathcal{B}_Z$. Hence, M is the unique minimizer of $F(Z)$.

5. We show uniqueness of the mean partition. Let $\mathcal{S}_n = (X_1, \dots, X_n)$ be a sample of n partitions. Let $\mathcal{S}_{\{n\}} = \{X_1, \dots, X_n\}$ denote the set of partitions induced by \mathcal{S}_n . We define a probability measure Q_n as a probability mass function of the form

$$Q_n(X) = \frac{1}{n} \sum_{i=1}^n \delta_{X, X_i}$$

for all $X \in \mathcal{S}_Q$, where $\delta_{X,Y}$ is the Kronecker delta that gives one when $X = Y$ agree and zero, otherwise. We use Q_n as probability measure and \mathcal{S}_Q as set containing the support of Q_n . Then the assertion follows from Part 1–4 of this proof.

B.2. Proof of Prop. 3.2

1. The first assertion holds for symmetric partitions Z , because $\alpha_Z = 0$ and therefore $\alpha_Z/4 \leq \rho_Z$. We assume that Z is asymmetric. The group Π is a discontinuous group acting isometrically on \mathcal{X} . The isotropy group

$$\Pi_Z = \{\mathbf{P} \in \Pi : \mathbf{P}Z = Z\}$$

is trivial for any representation $Z \in Z$. Since $\mathcal{P} \cong \mathcal{X}/\Pi$, we have a bijective isometry

$$\phi : \mathcal{B}(Z, \rho) \longrightarrow \mathcal{B}(Z, \rho), \quad \mathbf{X} \mapsto \pi(\mathbf{X}).$$

for all $0 < \rho \leq \alpha_Z/4$ by [17], Theorem 13.1.1. Setting $\psi = \phi^{-1}$ we find that $\mathcal{B}(Z, \rho)$ is a homogeneous ball. This shows $\alpha_Z/4 \leq \rho_Z$.

2. We show the second assertion. From Part 1 of this proof follows $\alpha_Z > 0 \Rightarrow \rho_Z > 0$. We show the opposite direction. Let $\rho_Z > 0$. We assume that $\alpha_Z = 0$. Suppose that $Z \in Z$ is a representation. Then there is a permutation matrix $\mathbf{P} \in \Pi^*$ such that $Z = \mathbf{P}Z$. Consider the ball $\mathcal{B}_\varepsilon = \mathcal{B}(Z, \varepsilon)$ for $\varepsilon > 0$. Suppose that $\mathbf{X} \in \mathcal{B}_\varepsilon$ is an element representing an asymmetric partition. Such an element exists according to Prop. 3.3(1). Then $\mathbf{X} \neq \mathbf{P}\mathbf{X}$ and $\mathbf{X}, \mathbf{P}\mathbf{X} \in \mathcal{B}_\varepsilon$. This shows that there is no bijective mapping between $\mathcal{B}(Z, \varepsilon)$ and \mathcal{B}_ε for any $\varepsilon > 0$. This contradicts our assumption that $\rho_Z > 0$. Hence, we have $\alpha_Z > 0$.

B.3. Proof of Prop. 3.3

1. To prove the first assertion, it is sufficient to show that the set of asymmetric partitions forms an open and dense subset in \mathcal{P} . The projection $\pi : \mathcal{X} \rightarrow \mathcal{X}/\Pi$ is open and surjective. Then the image $\pi(\mathcal{U})$ of an open and dense subset $\mathcal{U} \subseteq \mathcal{X}$ is open and dense in \mathcal{X}/Π . To see this observe that from $\pi(\mathcal{X}) = \pi(\overline{\mathcal{U}})$ and surjectivity of π follows $\pi(\overline{\mathcal{U}}) = \mathcal{X}/\Pi$. From $\mathcal{X}/\Pi = \pi(\overline{\mathcal{U}}) \subseteq \overline{\pi(\mathcal{U})}$ follows that $\pi(\mathcal{U})$ is open and dense in \mathcal{X}/Π .

Now let $\mathbf{Z} \in \mathcal{X}$ be a representation of an asymmetric partition $Z \in \mathcal{P}$. Suppose that $\mu : \mathcal{P} \rightarrow \mathcal{D}_{\mathbf{Z}}$ is a cross section. From Prop. A.3 follows that $\mu(\mathcal{P}) \subset \mathcal{D}_{\mathbf{Z}}$ is a fundamental set containing the open set $\mathcal{D}_{\mathbf{Z}}^{\circ}$. Since $\mathcal{D}_{\mathbf{Z}}^{\circ}$ is open and dense in $\mathcal{D}_{\mathbf{Z}}$, we find that $\pi(\mathcal{D}_{\mathbf{Z}}^{\circ})$ is open and dense in \mathcal{P} . This shows that almost all partitions are asymmetric.

2. We show the second assertion. Let $Z \in \mathcal{P}$ be an asymmetric partition. Suppose that $\mathbf{Z} \in Z$ is a representation matrix with rows $\mathbf{z}_1, \dots, \mathbf{z}_{\ell}$. We assume that Z has two identical clusters. Then there are two distinct indices $1 \leq p < q \leq \ell$ such that $\mathbf{z}_p = \mathbf{z}_q$. Let $\mathbf{P} \in \Pi^*$ be the permutation matrix that swaps rows p and q . Then we have $\alpha_Z \leq \|\mathbf{Z} - \mathbf{PZ}\| = 0$. This contradicts our assumption that Z is asymmetric. Hence, the clusters of Z are mutually distinct.

Next, we assume that all clusters of Z are mutually distinct. Suppose that Z is asymmetric. Then there is a representation $\mathbf{Z} \in Z$ and a permutation matrix $\mathbf{P} \in \Pi^*$ such that $\alpha_Z = \|\mathbf{Z} - \mathbf{PZ}\|$. As stated in Section A.1, we can express \mathbf{P} as a minimal product of $t > 0$ transpositions. Hence, representation \mathbf{Z} has at least one pair of identical rows. This contradicts our assumption that the clusters of Z are mutually distinct showing that partition Z is asymmetric.

B.4. Proof of Prop. 3.5

The first three parts of this proof prepare the proofs of the assertions shown in Part 4–6.

1. Let $\mathbf{P} \in \Pi^*$ be a transposition that permutes rows $p < q$. Then we have

$$\|\mathbf{Z} - \mathbf{PZ}\|^2 = \|\mathbf{z}_p - \mathbf{z}_q\|^2 + \|\mathbf{z}_q - \mathbf{z}_p\|^2 = 2\|\mathbf{z}_p - \mathbf{z}_q\|^2.$$

This gives $\|\mathbf{Z} - \mathbf{PZ}\| = \sqrt{2}\|\mathbf{z}_p - \mathbf{z}_q\|$.

2. Let $\mathbf{P}, \mathbf{Q} \in \Pi^*$ be two different transpositions. Suppose that \mathbf{P} permutes rows $p < q$ and \mathbf{Q} permutes rows $r < s$ such that either (i) $\{p, q\} \cap \{r, s\} = \emptyset$ or (ii) $q = r$. For case (i), we have

$$\|\mathbf{Z} - \mathbf{PQZ}\|^2 = 2\|\mathbf{z}_p - \mathbf{z}_q\|^2 + 2\|\mathbf{z}_r - \mathbf{z}_s\|^2$$

according to the first part of this proof. This implies $\|\mathbf{Z} - \mathbf{PZ}\| \leq \|\mathbf{Z} - \mathbf{PQZ}\|$ and $\|\mathbf{Z} - \mathbf{QZ}\| \leq \|\mathbf{Z} - \mathbf{PQZ}\|$.

3. As stated in Section A.1, we can write $\mathbf{P} \in \Pi^*$ as a matrix product $\mathbf{P} = \mathbf{Q}_1 \cdots \mathbf{Q}_t$ of transpositions $\mathbf{Q}_i \in \Pi$ with minimum number $t > 0$ of factors. From the second part of this proof follows

$$\|\mathbf{Z} - \mathbf{Q}_i \mathbf{Z}\| \leq \|\mathbf{Z} - \mathbf{Q}_1 \mathbf{Q}_2 \cdots \mathbf{Q}_i \mathbf{Z}\| = \|\mathbf{Z} - \mathbf{PZ}\|$$

for all $i \in \{1, \dots, t\}$. This shows that it is sufficient to restrict to transpositions for determining the degree of asymmetry of a partition.

4. Let Π_τ denote the subset of all transpositions. From part 1 and 3 of this proof follows

$$\alpha_Z = \min \{ \|\mathbf{Z} - \mathbf{P}\mathbf{Z}\| : \mathbf{P} \in \Pi_\tau \} = \min \left\{ \sqrt{2} \|\mathbf{z}_p - \mathbf{z}_q\| : 1 \leq p < q \leq \ell \right\}.$$

This shows the first assertion.

5. The second assertion assumes that Z is a hard partition. Then the elements of \mathbf{Z} take binary values from $\{0, 1\}$ such that $\mathbf{z}_p^T \mathbf{z}_q = 0$ for all $1 \leq p < q \leq \ell$. Moreover, $n_p = \mathbf{z}_k^T \mathbf{z}_k$ is the size of the k -th cluster. Then we have

$$\|\mathbf{z}_p - \mathbf{z}_q\|^2 = \mathbf{z}_p^T \mathbf{z}_p - 2\mathbf{z}_p^T \mathbf{z}_q + \mathbf{z}_q^T \mathbf{z}_q = n_p + n_q.$$

From the fourth part of this proof follows

$$\alpha_Z = \min \left\{ \sqrt{2(n_p + n_q)} : 1 \leq p < q \leq \ell \right\}.$$

This implies the second assertion.

6. The second assertion assumes that Z is an asymmetric hard partition. We first prove the lower bound of α_Z . From Prop. 3.3 follows that Z has at most one empty cluster. Then $m_1 = 0$ and $m_2 > 0$. Using part 5 of this proof, we obtain $\sqrt{2} \leq \sqrt{2m_2} = \alpha_Z$. Next, we show the upper bound of α_Z . From the strong form of the pigeonhole principle follows that there is a cluster with at least $\mu_1 = \lceil m/\ell \rceil$ elements. Let $m' = m - \mu_1$ be the number of remaining elements. Then again applying the pigeonhole principle gives a cluster with at least $\mu_2 = \lceil m' / (\ell - 1) \rceil$ elements. From $\mu_1 \geq m/\ell$ follows

$$\mu_2 = \frac{m - \mu_1}{\ell - 1} \leq \frac{m - m/\ell}{\ell - 1} = \frac{m}{\ell} \leq \left\lceil \frac{m}{\ell} \right\rceil.$$

We can bound the cardinality of the two smallest clusters by

$$m_1 + m_2 \leq \mu_1 + \mu_2 \leq 2 \left\lceil \frac{m}{\ell} \right\rceil.$$

Using part 5 of this proof shows the third assertion.

B.5. Proof of Equation (1)

We have

$$\begin{aligned} I_{n,k} &= \frac{1}{n^2} \sum_{i,j} \Delta_k(X_i, X_j) \\ &\geq \frac{1}{n^2} \sum_{i,j} \Delta_k(M_k, X_j) \\ &= \frac{1}{n} \sum_j \Delta_k(M_k, X_j) \\ &= F_{n,k}(M_k). \end{aligned}$$

The inequality in the second line holds, because M_k is a mean or medoid. This shows the assertion.

References

- [1] A. Bhattacharya and R. Bhattacharya. *Nonparametric Inference on Manifolds with Applications to Shape Spaces*. Cambridge University Press, 2012.
- [2] G. E. Bredon. *Introduction to Compact Transformation Groups*. Elsevier, 1972.
- [3] E. Dimitriadou, A. Weingessel, and K. Hornik. A Combination Scheme for Fuzzy Clustering. *Advances in Soft Computing*, 2002.
- [4] C. Domeniconi and M. Al-Razgan. Weighted cluster ensembles: Methods and analysis. *ACM Transactions on Knowledge Discovery from Data*, 2(4):1–40, 2009.
- [5] X.Z. Fern and W. Lin. Cluster ensemble selection. *Statistical Analysis and Data Mining*, 1(3): 128–141, 2008.
- [6] V. Filkov and S. Skiena. Integrating microarray data by consensus clustering. *International Journal on Artificial Intelligence Tools*, 13(4):863–880, 2004.
- [7] L. Franek and X. Jiang. Ensemble clustering by means of clustering embedding in vector spaces. *Pattern Recognition*, 47(2):833–842, 2014.
- [8] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’institut Henri Poincaré*, 215–310, 1948.
- [9] R. Ghaemi, N. Sulaiman, H. Ibrahim, and N. Mustapha. A Survey: Clustering Ensembles Techniques. *Proceedings of World Academy of Science, Engineering and Technology*, 38:644–657, 2009.
- [10] A. Gionis, H. Mannila, and P. Tsaparas. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data*, 1(1):341–352, 2007.
- [11] A.K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8): 651–666, 2010.
- [12] B.J. Jain. Geometry of Graph Edit Distance Spaces. *arXiv: 1505.08071*, 2015.
- [13] B.J. Jain. Asymptotic Behavior of Mean Partitions in Consensus Clustering. *arXiv:1512.06061*, 2015
- [14] T. Li, C. Ding and M.I. Jordan. Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization. *IEEE International Conference on Data Mining*, 2007.
- [15] M. Lichman. *UCI Machine Learning Repository*, [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
- [16] U. von Luxburg. *Clustering stability: An overview*. Now Publishers Inc., 2010.
- [17] J.G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Springer, 2006.
- [18] A. Strehl and J. Ghosh. Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.
- [19] A.P. Topchy, M.H. Law, A.K. Jain, and A. Fred. Analysis of consensus partition in cluster ensemble. *IEEE International Conference on Data Mining*, 2004.
- [20] A.P. Topchy, A.K. Jain, and W. Punch. Clustering ensembles: Models of consensus and weak partitions. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 27(12):1866–1881, 2005.
- [21] S. Vega-Pons, J. Correa-Morris and J. Ruiz-Shulcloper. Weighted partition consensus via kernels. *Pattern Recognition*, 43(8):2712–2724, 2010.
- [22] S. Vega-Pons and J. Ruiz-Shulcloper. A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03), 337–372, 2011.
- [23] Z.H. Zhou. *Ensemble methods: foundations and algorithms*. CRC Press, 2012.